

Polynomial Chaos-based Bayesian Inference of K-Profile Parametrization in a General Circulation Model of the Tropical Pacific

Ihab Sraj*

King Abdullah University for Science and Technology, Thuwal, Saudi Arabia

Sarah E. Zedler

University of Texas at Austin, Austin, TX, USA

Omar M. Knio

*Duke University, Durham, NC, USA; King Abdullah University for Science and Technology,
Thuwal, Saudi Arabia*

Charles S. Jackson

University of Texas at Austin, Austin, TX, USA

Ibrahim Hoteit

King Abdullah University of Science and Technology, Thuwal, Saudi Arabia

**Corresponding author address:* Ihab Sraj, Division of Physical Sciences and Engineering, King Abdullah University for Science and Technology, Thuwal, Saudi Arabia.

E-mail: ihab.sraj@kaust.edu.sa

ABSTRACT

The authors present a Polynomial Chaos (PC)-based Bayesian inference method for quantifying the uncertainties of the K-Profile Parametrization (KPP) within the MIT General Circulation Model (MITGCM) of the tropical pacific. The inference of the uncertain parameters is based on a Markov Chain Monte Carlo (MCMC) scheme that utilizes a newly formulated test statistic taking into account the different components representing the structures of turbulent mixing on both daily and seasonal timescales in addition to the data quality, and filters for the effects of parameter perturbations over those due to changes in the wind. To avoid the prohibitive computational cost of integrating the MITGCM model at each MCMC iteration, we build a surrogate model for the test statistic using the PC method. To filter out the noise in the model predictions and avoid related convergence issues, we resort to a Basis-Pursuit-DeNoising (BPDN) compressed sensing approach to determine the PC coefficients of a representative surrogate model. The PC surrogate is then used to evaluate the test statistic in the MCMC step for sampling the posterior of the uncertain parameters. Results of the posteriors indicate good agreement with the default values for two parameters of the KPP model namely the critical bulk and gradient Richardson numbers; while the posteriors of the remaining parameters were barely informative.

1. Introduction

The present work seeks to calibrate the model parameters of the K-Profile Parameterization (KPP) model (Large et al. 1994a, 1997) as implemented in the ocean model MIT General Circulation Model (MITGCM) (Ferreira and Marshall 2006; Marshall et al. 1997a) of the tropical Pacific. The KPP model relies on a number of parameters whose default values are set based on a combination of theory, laboratory experiments, and atmospheric/oceanic boundary layer observations (Large et al. 1994a, 1997). Our goal here is to quantify the uncertainties in these parameters where the ocean resolves the chaotic behavior of fluid dynamic models. The chaos in the response of the deterministic MITGCM model to the perturbation of these parameters leads to internal noise that in turn results in low signal to noise challenges as will be discussed in more detail below.

An inverse modelling approach is adopted for the objective stated above in which a set of temperature, salinity and horizontal current measurements are used to estimate the KPP parameters. Specifically, we employ a Bayesian approach to inverse problems that provides complete posterior statistics and not just a single value for the quantity of interest (Tarantola 2005). Traditionally, local model-data misfit of short-term turbulent mixing events are used to construct a cost function and then Bayesian inference is employed for the estimation of the uncertain parameters (Sivia 2006). Here, instead, we use a test statistic for KPP parameters' estimation that was introduced in Wagman et al. (2014) and in Zedler et al. (in revision). This statistic seeks to formulate a total cost function of different components representing the structures of turbulent mixing on both daily and seasonal timescales, takes data quality into account, and filters for the effects of parameter perturbations over those due to changes in the wind. At both timescales, the model and data are filtered before taking differences to capture the time integrated response, rather than individual mixing events.

The end result of the Bayesian inference formulation is a multi-dimensional posterior that can be directly sampled via Markov Chain Monte Carlo (MCMC). This, however, requires a prohibitive number of simulations of the forward model, one for every proposed set of parameters of the Markov chain (Malinverno 2002). This practice renders Bayesian methods computationally prohibitive for large-scale models such as the MITGCM where one model evaluation takes 22 hours in computing time using 256 processors. To overcome this issue, we construct a surrogate model that approximates the forward model and can be used in the sampling MCMC. More precisely, we use the Polynomial Chaos (PC) method to construct the surrogate model from an ensemble of MITGCM model runs (Marzouk et al. 2007; Marzouk and Najm 2009). This approach further offers additional advantages such as computing model output sensitivities and additional statistics (Le Maître and Knio 2010).

The PC method has been extensively investigated in the literature, and its suitability for large-scale models has been recently demonstrated in various settings. Alexanderian et al. (2012) implemented a sparse spectral projection PC approach to propagate parametric uncertainties of three KPP parameters in addition to the wind drag coefficient during a hurricane event. The study demonstrated the possibility of building a representative surrogate model for a realistic ocean model; however the inverse problem was not tackled. Winokur et al. (2013) followed up on Alexanderian et al. (2012) work and implemented an adaptive strategy to design sparse ensembles of oceanic simulations for the purpose of constructing a PC surrogate with even less computational effort. Sraj et al. (2013b) extended the previous work and combined a spectral projection PC approach with Bayesian inference to estimate parametrized wind drag coefficient using temperature data collected during a typhoon event. The same problem was also solved using gradient based search method (Sraj et al. 2013a). Mattern et al. (2012) have similarly exploited the virtues of such polynomial expansions for examining the response of ecosystem models to finite perturbations of

their uncertain parameters. Tsunami (Sraj et al. 2014; Ge and Cheung 2011), climate (Olson et al. 2012) and subsurface flow modeling (Elsheikh et al. 2014) were also investigated using similar PC approaches.

What is common in the aforementioned PC applications is that the processes studied occurred over short timescales of few days, so that internal noise in the model was small. This enabled a successful PC expansion construction using traditional spectral projection (Sraj et al. 2013b; Reagan et al. 2003; Alexanderian et al. 2012). In this study, however, the major hurdle of constructing a PC surrogate model was the internal noise present in the MITGCM model due to the perturbation of the chosen KPP parameters, which was amplified over time by non-linear interactions. As explained in Section 4b(1), a spectral projection technique failed to construct a PC expansion that faithfully represent the model. Instead, we resort to a compressed sensing technique namely the Basis-Pursuit-DeNoising (BPDN) method (Peng et al. 2014) to determine the PC expansion coefficients. This technique first seeks to estimate the noise in the model output, filter it out and then solve an optimization problem assuming sparsity in the PC expansion to determine the non-zero PC coefficients efficiently. The BPDN method offers an additional advantage that a smaller number of model runs compared to the spectral projection method is required to determine the PC coefficients as shown in Section 4c. BPDN was recently employed to build a proxy model for an integral oil-gas plume model (Wang et al. 2015) and for an ocean model with initial and wind forcing uncertainties (Li et al. 2015). In the former, the model output was noisy due to the iterative solver used in the double-plume calculation (Socolofsky et al. 2008) and BPDN proved to be successful in filtering the noise and thus building a representative PC model. In the latter, no noise was present in the model output, however, the model was unable to produce realistic simulations for pre-specified sets of parameters as required by the spectral projection method. BPDN was therefore used as an alternative approach as it does not have this requirement, a random sample of

model parameters can be modeled instead (Doostan and Owhadi 2011) . To our best knowledge, BDPN has not previously been applied to a noisy, large scale ocean model.

The structure of the paper is as follows: Section 2 describes the MITGCM, presents the choice of the uncertain parameters and describes the observation and cost function used in the estimation process. Section 3 introduces the Bayesian inference and its application to our specific problem. Section 4 discusses the PC method and presents several error studies to show the convergence of the constructed PC expansion. Section 5 presents the results of the inference of KPP parameters and Section 6 summarizes our findings.

2. Model, Uncertain parameters and Observations

a. MITcgm Model

The MITGCM employed in this work is based on the primitive Navier Stokes equations implemented in spherical coordinates with an implicit non-linear free surface. The MITGCM implements the K-Profile Parameterization (KPP) turbulent mixing scheme (Adcroft 1995; Marshall et al. 1997a,b; Large et al. 1994b) (see below) in a regional configuration based on that of Hoteit et al. (2008) and Hoteit et al. (2010) for the simulation of oceanic flow. In particular, the domain chosen covers the region with latitudes from 26°S to 30°N and longitudes from 104°E to 70°W (Figure 1). The time period of our model simulation is 2004-2007. The initial and lateral boundary conditions are provided by the Ocean Comprehensible Atlas (OCCA) reanalysis that was developed for the 2004-2007 time period using data assimilation in the MITGCM of available temperature and salinity ocean data sets (Forget 2010). The lateral boundary conditions for our model are implemented with a sponge layer (with a thickness of 9 grid cells, and inner and outer boundary relaxation timescales of 20 and 1 days, respectively). For the lateral boundary condi-

tions, the OCCA data assimilation product is interpolated at the model resolution of $1/3^\circ$ and with a time step of one day. Therefore, at the initial timestep, the boundary temperature and salinity conditions are approximately in equilibrium with the interior fields. Once our higher resolution simulation starts, the velocity field quickly adjusts to the pressure gradient forces and establishes a realistic Equatorial circulation. We note that our model runs on 256 processors and takes about 22 hours for a single simulation. As described below in Section 4, we needed to run the model 903 times, which required a total of about 5.5 million compute hours.

b. KPP model

In the ocean, turbulent mixing can ensue when there is net heat released to the atmosphere at the sea surface (i.e. at night), producing gravitationally unstable density inversions (convective mixing) and when there is sufficient vorticity-producing (in the x-z plane) vertical shear in the horizontal currents to overturn a nominally stratified water column (shear-driven or Kelvin-Helmholz instability induced mixing). In general terms, the intensity of convective and shear-driven mixing depend on local water column properties and surface forcing conditions. Theoretically, the most vigorous turbulent mixing should occur when a weakly unstably stratified, strongly sheared flow is forced with strong winds and convection (i.e. at night). By contrast, the flow is most likely to be laminar when a strongly stably stratified, weakly sheared flow is forced with weak winds and large net heat going into the ocean (i.e. during the day). The KPP embodies these basic relationships, by making the intensity of mixing a function of locally diagnosed properties of the water column that relate to the amenability to turbulent mixing (such as the bulk and gradient Richardson numbers) as well as the non-local surface wind stress and net heat flux forcing (through forcing parameters such as the friction velocity of wind and the Monin-Obukhov length). In the KPP, mixing is more

intense under unstable convective surface forcing conditions. Readers interested in the details of the KPP are referred to Large et al. (1994b) and Large and Gent (1999).

The KPP generates depth profiles of two quantities that are relevant for turbulent mixing, the eddy diffusivity/viscosity and a non-local term. The eddy diffusivity and viscosity at depth z can be thought of as a scale of the intensity of the turbulent mixing there, with larger values indicating more vigorous turbulence. The role of the non-local term is to enhance turbulent fluxes of temperature and salinity (but not the horizontal velocity components) under convective forcing conditions.

There are nine parameters in the KPP that pertain to convective or shear-driven mixing. A list of the five uncertain parameters in the KPP model under investigation in this work is presented in Table 1 as well as minimum and maximum values for the uniform prior assumed for each. The default values in MITGCM are also indicated in the table. The critical bulk and gradient Richardson (Ri_c and Ri_g , respectively) numbers relate directly to local water column shear/stratification properties, with larger values generally making turbulent mixing more intense (for the same water-column). Convective mixing parameters $\phi_{s,unst}$ and $\phi_{m,unst}$ depend directly on a combination of surface forcing and local water column shear/stratification considerations and are zero under stable forcing conditions (during the day). Increasing their value generally makes convection more vigorous (given the same water column properties and surface forcing). The non-local convective mixing parameter C^* is proportional to the non-local convective term, so increasing it will make the turbulent fluxes for temperature and salinity larger.

c. Observations and test statistic

The observational data for our experiment come from the TOGA-TAO mooring array for the November 2003–November 2007 time period. The array consists of 77 moorings, shown in Fig-

ure 1, centered on the Equator that span the width of the tropical Pacific in the east-west direction, in the latitude range from 8°S to 8°N (McPhaden et al. 1998). The data used included measurements of temperature, salinity and horizontal current components.

The test statistic used in this paper has components that operate on daily and seasonal timescales. At daily timescales, it measures the model’s ability to reproduce the observed relationship between wind forcing and subsequent lowering of the sea surface temperature that results from shear-driven mixing. Prior to calculating the correlation between those quantities, the model and data are filtered to remove the diurnal cycle. At seasonal timescales, the test statistic measures the ability of the model to reproduce patterns of sensitivity in the ocean state that would result from perturbing KPP parameters (as diagnosed from an ensemble of single KPP parameter perturbation experiments). The patterns of sensitivity are determined separately for temperature, salinity, east velocity, and north velocity as extracted on the TOGA/TAO sensor array. For more details of what and how data was used for the calculation of the test statistic, the reader is referred to Wagman et al. (2014) and Zedler et al. (in revision).

3. Bayesian Inference

Bayesian inference is a statistical approach to inverse problems (Sivia 2006) that has recently gained great interest in different applications, including ocean (Alexanderian et al. 2011; Zedler et al. 2012; Sraj et al. 2013b), tsunami (Sraj et al. 2014) climate (Olson et al. 2012) and geophysical (Malinverno 2002) modeling. We briefly review this approach below.

a. Formalism

Let $d = (d_1, \dots, d_n)^T$ be a vector of observation data and $p = (p_1, \dots, p_m)^T$ be a vector of model parameters. We consider a forward model G that predicts the data as function of the parameters such that:

$$d \approx G(p). \quad (1)$$

Let $\pi(p)$ be the prior probability distribution of p (representing any *a priori* information on p), $L(d|p)$ the likelihood function (the probability of obtaining d given p), and $\pi(p|d)$ the posterior probability distribution of p (the probability of occurrence of p given d). In this case, the Bayes' rule governs this formulation:

$$\pi(p|d) \propto L(d|p) \pi(p). \quad (2)$$

The expression of the likelihood function depends on the assumptions made on the errors ε (discrepancies) between the model and observations ($\varepsilon = d - G(p)$). It is often assumed that the errors ε_i are independent and normally distributed with a covariance Σ . Traditionally, a metric $E(p)$, called the cost function, is constructed from the sum over squared errors normalized by estimates of the variances:¹

$$E(p) = \frac{1}{2} \varepsilon^T \Sigma^{-1} \varepsilon. \quad (3)$$

In this case the likelihood function can be written as:

$$L(d|p) = \frac{1}{(2\pi)^{\frac{ke}{2}} |\Sigma|^{\frac{1}{2}}} \exp(-E(p)) \quad (4)$$

¹We note that in this work a new test statistic is adopted to construct metric E as described in Section 2c and in Zedler et al. (in revision).

and the joint posterior in Equation 2 is then expressed as:

$$\pi(p|d) \propto \frac{1}{(2\pi)^{\frac{ke}{2}} |\Sigma|^{\frac{1}{2}}} \exp(-E(p)) \prod_{i=1}^m \pi(p_i). \quad (5)$$

The prior of the parameters p_i is assumed non-informative, i.e. uniform distribution such that $\pi(p_i) = \frac{1}{b_i - a_i}$ where a_i and b_i are the bounds of the prior indicated in Table 1 and ke is the effective degree of freedom.

To account for missing information about off-diagonal coefficients of the covariance matrix Σ , the test statistic E can be re-scaled by a parameter S . Incorporating a scaling parameter S is common in statistical inference as a way of scaling model fit to data given the level of agreement of the model with the data (Jackson and Huerta 2015; Jackson et al. 2004). The scaling parameter S is considered as an additional parameter in the Bayesian problem. and added to the test statistic E in the above equations where the joint posterior becomes as follows:

$$\pi(p, S|d) \propto \frac{S^{\frac{ke}{2}}}{(2\pi)^{\frac{ke}{2}} |\Sigma|^{\frac{1}{2}}} \exp(-SE(p)) \pi(S) \prod_{i=1}^m \pi(p_i). \quad (6)$$

The scaling parameter is treated as a hyper-parameter and its prior $\pi(S)$ is taken as a Gamma function (Wang and Zabarar 2005; Gelman et al. 2004; Gelman 2006) that depends on two constants α and β as follows:

$$\pi(S) = \begin{cases} \frac{\beta^\alpha}{\Gamma(\alpha)} S^{\alpha-1} \exp(-\beta S), & S > 0 \\ 0, & \text{otherwise.} \end{cases} \quad (7)$$

In our case we choose $\alpha = 18.18$ and $\beta = 72.02$. The prior of S thus has normal-like distribution with mean value $\frac{\alpha}{\beta} = 0.252$ and variance $\frac{\alpha}{\beta^2} = 0.003$. We determined the effective degrees of freedom for our integrated test statistic to be $ke = 17$.

b. Sampling method

Inferring the KPP parameters amounts to sampling the posterior in Equation 6. In general, when the space of the unknown parameters is multidimensional, a suitable computational strategy is the Markov Chain Monte Carlo (MCMC) method, yet MCMC requires a high number of sampling iterations. In our case, sampling the posterior however requires sampling the MITGCM model, which is computationally prohibitive. Thus we seek to build a surrogate model for the Quantity of Interest (QoI) as described in the following section, and use the random walk Metropolis MCMC algorithm (Roberts and Rosenthal 2009; Haario et al. 2001) to accurately and efficiently sample the posterior. Since the scaling parameter S is included in our test statistic as a scalar correction to the data covariance matrix, it is also included as a hyper-parameter to be estimated in addition to the model parameters p . We assume the priors for p and S are independent. This implies that for each MCMC step we can use Gibbs sampling to iteratively generate a value of p conditional on S and a value of S conditional on p as follows:

1. We simulate p conditional on S , apply sampling algorithm for p but for just one iteration.
2. We simulate S conditional on p ; for the informative gamma distribution, we have:

$$\pi(S|p, d) \propto S^{\frac{k_e}{2} + \alpha - 1} \exp(-S[E(p) + \beta]), \quad (8)$$

which results in a gamma distribution of parameters $\frac{k_e}{2} + \alpha$ and $E(p) + \beta$.

The two steps are repeated until convergence.

4. Accelerating Bayesian Inference

To reduce the cost of sampling the posterior, we rely on a surrogate model of the QoI that requires a much smaller ensemble of model runs (Malinverno 2002; Marzouk and Najm 2009). Here, we

rely on Polynomial Chaos expansions for representing the QoIs, which, in addition can efficiently provide statistical properties, such as the mean, variance and sensitivities (Crestaux et al. 2009).

Due to the complexity of the MITGCM, constructing a surrogate for the different model outputs is not feasible. Instead, we construct a single surrogate for the test statistic E which is the QoI in this case. This test statistic E is computed from the outputs of the model runs required for the construction of the surrogate as explained below. This practice simplifies the PC calculation where only one surrogate model would be constructed that can be sampled directly in the posterior of Equation (6).

a. Polynomial Chaos

Polynomial Chaos (PC) is a probabilistic methodology that expresses the dependencies of QoI on the uncertain model inputs as a truncated polynomial expansion (Ghanem and Spanos 1991; Villegas et al. 2012; Lin and Karniadakis 2009; Xiu and Tartakovsky 2004). The PC method is briefly described below; for more details the reader is referred to Le Maître and Knio (2010).

We show here the process of constructing a PC surrogate for the test statistic E . To this end, we denote by $\xi = (\xi_1, \dots, \xi_m)$ the canonical vector of random variables that parametrize the uncertain inputs i.e. the KPP parameters. In the case of uniform distributions the canonical vectors are calculated as follows $\xi_i = \frac{2p_i - (a_i + b_i)}{(a_i - b_i)}$ (Le Maître and Knio 2010). PC expresses the dependencies of E on the uncertain input variables ξ as a truncated expansion of the following form:

$$E(\xi) \approx \sum_{k=0}^R e_k \psi_k(\xi), \quad (9)$$

where e_k are the polynomial coefficients, and $\psi_k(\xi)$ are elements of an orthogonal basis of an underlying probability space. The total number of terms in the truncated PC expansion is $R + 1 =$

$\frac{(m+r)!}{m! r!}$ where m is the number of stochastic dimensions and r is the highest order polynomial retained.

The choice of the basis is dictated by the probability density function $\rho(\xi)$ of the stochastic vector ξ , which appears as a weight function in the probability space's inner product:

$$\langle \psi_i, \psi_j \rangle = \int \psi_i(\xi) \psi_j(\xi) \rho(\xi) d\xi = \delta_{ij} \langle \psi_i^2 \rangle, \quad (10)$$

where δ_{ij} is the Kronecker delta. For uniform distributions (our case), the basis functions are scaled Legendre polynomials. For multi-dimensional problems the basis functions are tensor products of 1D basis functions (Le Maître and Knio 2010).

b. Determination of PC coefficients

Various methods have been proposed for the determination of the PC coefficients e_k . They can be classified into Non-intrusive and Galerkin methods. Non-intrusive methods rely on an ensemble of deterministic model evaluations of $E(\xi)$, for particular realizations of ξ selected either at random or deterministically. Non-Intrusive methods include Non-Intrusive Spectral and Pseudo-Spectral Projection (Reagan et al. 2003; Constantine et al. 2012; Conrad and Marzouk 2013), Least-Square-Fit and regularized variants (Berveiller et al. 2006; Blatman and Sudret 2011; Peng et al. 2014), Collocation (interpolation) methods (Babuška et al. 2007; Xiu and Hesthaven 2005; Nobile et al. 2008), that are often combined with Sparse-Grid algorithms to reduce computational complexity. In the present paper, we adopt non-intrusive approaches that allow the use of the forward model as a black box with no code modifications required. PC expansion coefficients are determined based on a set of response simulations for a specified set of the uncertain parameters.

1) NON-INTRUSIVE SPECTRAL PROJECTION

We first applied on the traditional Non-Intrusive Spectral Projection (NISP) method that exploits the orthogonality of the basis and applies a Galerkin projection to find the PC expansion coefficients as follows:

$$e_k = \frac{\langle E, \psi_k \rangle}{\langle \psi_k, \psi_k \rangle} = \frac{1}{\langle \psi_k, \psi_k \rangle} \int E \psi_k(\xi) \rho(\xi) d\xi. \quad (11)$$

This orthogonal projection minimizes the L_2 error on the space spanned by the basis. The stochastic integrals are then approximated using a numerical quadrature to obtain:

$$\langle E, \psi_k \rangle \approx \sum_{q=1}^Q E(\xi_q) \psi_k(\xi_q) \omega_q, \quad (12)$$

where ξ_q and ω_q are multi-dimensional quadrature points and weights, respectively, and Q is the number of nodes in the multi-dimensional quadrature. The quadrature order should be commensurate with the truncation order, and should be high enough to avoid aliasing artifacts. The choice of quadrature rule is hence critical to the performance of the PC.

The computation of the e_k can be finally expressed as a matrix-vector product of the form:

$$e_k = \sum_{q=1}^Q \mathcal{P}_{kq} E(\xi_q), \quad \mathcal{P}_{kq} = \frac{\psi_k(\xi_q) \omega_q}{\langle \psi_k, \psi_k \rangle}, \quad (13)$$

where \mathcal{P}_{kq} is called the NISP projection matrix (can be pre-computed) and $E(\xi_q)$ is obtained from an ensemble of the deterministic model realizations with the uncertain parameters set at the quadrature values ξ_q .

In our present work, a quadrature was built based on Smolyak sparse nested grid (Petrus 2000, 2001, 2003; Gerstner and Griebel 2003; Smolyak 1963) to reduce the number of expensive deterministic MITGCM runs. For a PC expansion of order $r = 5$ (total number of terms in the truncated

PC expansion $R + 1 = 252$) and uncertain parameters $m = 5$, a total number of $Q = 903$ quadrature nodes were needed corresponding to Smolyak level 5. A projection of the quadrature nodes is shown in Figure 2 on two-dimensional plane.

We therefore ran MITGCM 903 times as per the quadrature and calculated the test statistic from the different model outputs. The test statistic corresponding to the sparse quadrature is shown in Figure 3 function of different parameters' spaces as indicated in each panel. The red dashed line in each panel represents the cases where the other parameters are set to the center of the corresponding priors i.e. $\xi_i = 0$. These figures also clearly show the uncertainty bounds in the test statistic due to the uncertainty in the input parameters. This is true for all five parameters.

The PC expansion coefficients are computed from the output of the 903 quadrature runs using Equation 13. Figure 4 plots the spectrum of the normalized PC coefficients, e_k/e_0 , in absolute value. The vertical lines separate the PC expansion terms into degrees $r = 0, \dots, 5$. The spectrum shows clearly that the PC suffers from convergence issues as the NISP-estimated PC coefficients do not decay with further increasing PC order but instead grow. This can be attributed to the presence of internal noise in the model that is not tolerated by the NISP method and thus over-fitting the model from the quadrature points with additional refinement of the PC order. To further asses this convergence issue, we show in Figure 5 the deterministic MITGCM realizations plotted on top of their PC expansion counterparts. The difference between the two sets of data confirms the inability of the NISP-estimated surrogate PC model to efficiently represent the QoI. In Figure 6, we show the test statistic E corresponding to 89 MITGCM model runs where we vary Ri_c only infinitesimally as indicated on the plot. The test statistic value is highly sensitive to infinitesimal perturbations of Ri_c . This likely results from internal noise in the model that is amplified over time, in part by non-linear interactions in the model.

As a conclusion, the construction of a converging PC expansion using the NISP method is not successful and the surrogate model is not representative of the QoI of MITGCM; therefore it can not be used for further analysis.

2) BASIS-PURSUIT DENOISING

In an attempt to find a suitable PC surrogate model for the test statistic E , we resort to a different approach that tolerates noise in the model but also that is non-intrusive. Instead of using spectral projection, we employ a recent technique that uses Compressed Sensing (CS) for polynomial representations which first estimates the noise in the model then determines the PC coefficients. Let $e = (e_0, \dots, e_R)$ be a vector of PC coefficients to be determined and let $E = (E(\xi_1), \dots, E(\xi_Q))$ be a vector of forward model evaluations function of sampled ξ_q . We also define Ψ as the matrix where each row corresponds to the row vector of $R + 1$ PC basis functions evaluated at the sampled ξ_q . CS solves the problem:

$$E = \Psi e^2 \quad (14)$$

by exploiting the approximate sparsity of the signal i.e. the vector of PC coefficients e that necessarily converge to zero. The sparsity is set by constraining the system and minimizing its energy i.e. its l_1 -norm. CS seeks a solution with minimum number of non-zero entries by solving the optimization problem:

$$\mathcal{O}_{1,\delta} \approx \left\{ \underset{e}{\operatorname{argmin}} \|e\|_1 : \|E - \Psi e\|_2 \leq \delta \right\} \quad (15)$$

where δ is the noise estimated in the signal.

²compare with Equation (9)

This l_1 -minimization problem is referred to as Basis-Pursuit (BP) when $\delta = 0$, and to Basis-Pursuit-Denoising (BPDN) when a noise δ in the system is assumed as proposed in Donoho (2006). We adapt the latter approach since we acknowledge the existence of noise in the predicted QoI. We note that in Equation (15) the constraint depends on selected sampled parameters ξ_q and their corresponding $E(\xi_q)$ and not on a general sample of ξ and $E(\xi)$. As a result, the coefficients e may be chosen to fit the input realizations, and not accurately approximate the model. To avoid this situation, we determine the noise δ by cross-validation as discussed in Peng et al. (2014). To solve $\mathcal{O}_{1,\delta}$ standard l_1 -minimization solvers may be used. In this work we use the MATLAB package SPGL1 (Berg and Friedlander 2007) based on the spectral projected gradient algorithm (van den Berg and Friedlander 2008).

Here, we applied the BPDN to determine the PC coefficients for the surrogate model of the test statistic. Instead of using Monte Carlo sampling method to generate realizations as in Peng et al. (2014) we take advantage of the previously simulated 903 MITGCM realizations and use them to solve the optimization problem. The resulting PC normalized coefficients spectrum e_k/e_o is plotted in Figure 7 (in absolute value) up to order $r = 5$ to assess the convergence of the PC expansion. The vertical lines indicate the PC terms for different polynomial order $r = 0, \dots, 5$. The spectrum shows clearly that the PC converges better with increasing PC order using BPDN as opposed to the NISP method.

To check the convergence of the PC expansion with PC order r , the PC surrogate is sampled to find the *pdf* of the test statistic E using different polynomial orders. The *pdfs* are shown in Figure 8 where we observe that as the PC order is increased, the *pdfs* get closer to each other, indicating robust estimates.

To confirm the ability of the constructed PC expansion to represent the QoI, we again show the MITGCM realizations plotted on top of their PC counterparts in Figure 9 (left). One interesting

observation is the internal noise in MITGCM that appears in the corresponding test statistic, while PC gives a smooth function for E . While PC filters out the internal noise in the QoI, it clearly captures the mean of the signal. Figure 9 (right) shows the same but in a scatter plot where MITGCM realizations are plotted against their PC counterparts. To quantify the agreement, we use another common error metric where we calculate the normalized relative error (NRE) between the 903 MITGCM simulations and their reconstruction using the built PC as follows:

$$NRE = \frac{\left(\sum_{q=1}^Q \left| E(\xi_q) - \sum_{k=0}^R e_k \psi_k(\xi_q) \right|^2 \right)^{1/2}}{\left(\sum_{q=1}^Q \left| E(\xi_q) \right|^2 \right)^{1/2}}, \quad (16)$$

The NRE is calculated and found to be $\sim 3\%$ which is acceptable, suggesting that BPDN is successful in constructing a surrogate that is able to accurately simulate the QoI.

For further validation of our surrogate model, an independent set of 954 MITGCM runs were conducted simultaneously by varying the same set of KPP uncertain parameters in a identical model setup and using the same test statistic. The sample is shown in Figure 10 as a 2-D projection of the $\xi_1 - \xi_2$ plane (representing Ri_c and Ri_g). The corresponding test statistic E is shown in Figure 11 as a function of the different parameters' spaces. The different plots in Figure 11 shows a functional trend mainly for Ri_c and Ri_g ; however, the variation in E suggests an internal noise in the model outputs in addition to responding to other parameters.

The PC expansion is calculated for the different sample points and is shown in Figure 12 (left) versus the MITGCM realization. The PC expansion appears to reproduce the mean of the deterministic model. Figure 9 (right) shows the same but in a scatter plot where MITGCM realizations are plotted against their PC counterparts. The NRE is calculated and found to be $\sim 4.2\%$, which

is also quite acceptable indicating that BPDN is successful in producing a surrogate that is able to accurately simulate the QoI.

c. PC sensitivity to ensemble size

Normally a number N of random samples is required to compute the PC coefficients using BPDN; this number is much less than the number of the unknowns i.e. the size of the PC expansion ($R + 1 = 252$) such that $N \ll R + 1$. In the above results, we instead used an ensemble of $Q = 903$ sparse quadrature MITGCM model runs corresponding to Smolyak level 5 (Figure 2) to compute the PC coefficients using BPDN due to their availability (upon using NISP method) where in this case $N = Q \gg R + 1$. Here, we explore the possibility of utilizing a small number of MITGCM realizations to build a faithful PC expansion surrogate. To create a smaller ensemble and avoid running new expensive MITGCM simulations, we consider lower levels of refinement of Smolyak quadrature shown in Figure 13 in the canonical vector space for levels 1, 2, 3 and 4 with total number of nodes 11, 51, 151 and 391, respectively. These levels are nested and therefore the nodes in each level is a subset of the higher levels. Thus, a small number of model runs N can be extracted from the original $Q = 903$ runs as per the Smolyak levels and used to construct different PC models using BPDN.

In Figure 14, we show the NRE of the original 903 Smolyak runs computed using PC models built using quadrature nodes from the different Smolyak levels. We observe that the error increases as N decreases; however the increase is not significant. In fact, even with level 2 (51 model runs) the NRE is around 5%. We note that in all the results below we used the PC expansion model constructed using the level 5 Smolyak runs with BPDN.

d. Statistical moments and sensitivity analysis

The identification of the inner product weight function with the probability distribution of ξ simplifies the calculations of the statistical moments of E . Noting that since $\psi_0(\xi)$ is a constant that is normalized so that $\langle \psi_0, \psi_0 \rangle = 1$, the mean and variance of E can be computed as:

$$\mu_E = \int E \rho(\xi) d\xi \approx \langle E, \psi_0 \rangle = e_0, \quad (17)$$

and

$$\sigma_E^2 = \int (E - \mu_E)^2 \rho(\xi) d\xi \approx \sum_{k=1}^R e_k^2 \langle \psi_k, \psi_k \rangle. \quad (18)$$

PC representations also enable efficient global sensitivity analysis that quantifies the contributions of different random input parameters to the variance in the output. This can be done by computing the so-called *total* sensitivity index T_i that measures the contribution of the i^{th} random input to total model variability by computing the fraction of the total variance due to all the terms in the PC expansion that involve ξ_i as follows:

$$T_i = \frac{\sum_{k \in K_i} e_k^2 \langle \psi_k^2 \rangle}{\sum_{k=1}^R e_k^2 \langle \psi_k^2 \rangle}, \quad K_i = \left\{ k \in \{1, \dots, R\} : \alpha_i^k > 0 \right\}, \quad (19)$$

where α^k is the multi-index associated with the k^{th} term in the PC expansion (Le Maître and Knio 2010; Crestaux et al. 2009; Sudret 2008).

The PC expansion is used to calculate statistical moments of E due to the uncertainty in the input parameters as indicated above. The mean of the test statistic was found to be 325.62 with a standard deviation of 30.97. The sensitivities were also computed and summarized in Table 2. It is

notable that the test statistic is most sensitive to Ri_c and then to Ri_g and $\phi_{m,unst}$ as clearly reflected in the corresponding total sensitivities, $T_1 - T_3$ respectively.

e. Response surfaces

In addition to the moments and sensitivities, the PC surrogate can be used to construct a response surface for E as a function of the uncertain input parameters. To this end, we sample the PC surrogate for different values of the canonical vector of random variables ξ within the prior range $[-1, +1]$ as illustrated in Figure 15. The different plots represent the response curves function of single uncertain parameter while the other parameters are set to $\xi = 0$. The plots show the strong dependance of E on Ri_c , consistent with the sensitivity results shown earlier. Similarly E shows some variations function of Ri_g and $\phi_{m,unst}$ but with a relatively less dependance compared to Ri_c . The remaining curves exhibit lines with a small slope, suggesting that E depends only mildly on $\phi_{s,unst}$ and C^* . We also show the 2D response surfaces for Ri_c versus Ri_g in Figure 16 (top), and Ri_c versus $\phi_{m,unst}$ in Figure 16 (bottom) (the other parameters are set to $\xi = 0$). The quadrature sample E_q are shown on top of the surfaces for comparison.

5. KPP Inference

Bayesian inference is now used to estimate the KPP parameters such that the likelihood based on the test statistic is maximized. To this end, a random-walk MCMC method is implemented to sample the posterior distributions (Roberts and Rosenthal 2009; Haario et al. 2001) (Equations 6 and 8) and consequently update the KPP parameters' distributions. This sampling requires tens of thousands of MITGCM runs that are prohibitively expensive as each MCMC sample requires an independent MITGCM realization. Instead, the surrogate model created using PC expansions

provides a computationally efficient alternative that requires only evaluating the PC expansion for different values of the canonical vector of random variables ξ .

All of the results presented here are based on 10^6 MCMC samples; we find negligible changes in the obtained posteriors of the KPP parameters with further iterations. Figure 17 shows the sample chains for the input parameters and scaling parameter S for different iterations of the MCMC algorithm. The different panels suggest well-mixed chains for all input parameters where the chains of all parameters appear to be concentrated in an area of the parameter prior range. The running mean plotted in Figure 18 is an indication of the convergence of the MCMC. For Ri_c , the chain appears to be concentrated in the lower end of the parameter range with values between 0.1 and 0.4 while for Ri_g , the chain appears to be concentrated in the upper end of the parameter range with values between 0.6 and 1.0. These values align well with what is commonly considered physically relevant for such parameters shown as horizontal lines in each panel. For $\phi_{m,unst}$ and $\phi_{s,unst}$, the chains appear to cover almost all the prior range. This is an indication of a non-informative posterior due to lack of data to infer those parameters. As for C^* , the chain is concentrated in the upper end of the parameter range between 10 and 15 as shown. Finally, the hyper-parameter S is well-mixed with values ranging between 0.03 and 0.12.

Next, the computed MCMC chains are used to determine the marginalized posterior distributions using Kernel Density Estimation (KDE) (Parzen 1962; Silverman 1986). The resulting marginalized posterior *pdfs* of the KPP are shown in Figure 19 in addition to the scaling parameter S . Note that the first 2×10^5 iterations, associated with the burn-in period, were discarded. As expected from the chains shown in Figure 17, the posterior *pdf* of Ri_c appears to be skewed to the right with an extended tail towards the higher Ri_c values. The pdf exhibits a well-defined peak but due to the skewness, we instead report the mean estimate calculated as ~ 0.23 ; it is close to the default MITGCM value (shown as a vertical line on the same pdf). For Ri_g , we observe a posterior that has

a well defined peak of 0.68; it matches the MITGCM default value. In contrast, the marginalized posterior *pdfs* of $\phi_{m,unst}$ and $\phi_{s,unst}$ appear to be fairly flat, and similar to the uniform prior; an indication that the observed data were not informative to refine our prior knowledge of those variables. Regarding C^* , the pdf indicates that the mean value is in the upper range of the prior. Finally the scaling parameter S exhibits a Gaussian like shape contrasted to the prior gamma distribution.

The posterior distributions are consistent with results of previous efforts to calibrate the KPP model (Large et al. 1994a; Large and Gent 1999). They show that the default critical bulk and gradient Richardson numbers and the non-local convective parameter C^* are within the range of acceptable parameter settings. In the absence of a detailed analysis of the role of $\phi_{m,unst}$ and $\phi_{s,unst}$ in our calibration, we can remark on possible explanations for their broad posterior distributions. It must be the case that either the eddy diffusivity and viscosity are weak functions of these parameters, or that there are negative feedbacks in the model that ultimately limit their influence on the ocean state. Since under equivalent local ocean conditions the functional dependence of eddy diffusivity on $\phi_{s,unst}$ or eddy viscosity on $\phi_{m,unst}$ implies variation in their values by a factor of 2-3 over the range in our priors, we believe that the latter explanation is more likely (Large et al., 1994, see Eq B1). The fact that the posterior distributions are in line with previous research is encouraging for the application of uncertainty quantification methods such as polynomial chaos to calibrate a model using observations. The question of whether data can be used to build more predictive models is still open.

6. Summary and conclusions

In this work, we presented a polynomial chaos-based method for the inference of five KPP parameter. below, we remark on the PC construction process and the KPP inference results.

The method relied on building a surrogate model for a newly developed test statistic instead of the model output. The advantage was to avoid building surrogate for several model outputs and for different time scales. The PC construction was implemented using two techniques. First a non-intrusive spectral projection method was used that required a quadrature of level 5 nodes ($Q = 903$ model runs). The computed PC expansion suffered from convergence issues due to the presence of internal noise in the predictions. This resulted in over-fitting of the data with increasing level of PC refinement. The resulting PC model was thus unsuitable and discarded. The second technique used was BPDN that tolerates noise in the model and assumes sparsity in the PC expansion, requiring a smaller number of model runs. This technique proved to be successful as it filters out the noise from the model in finding the PC coefficient that lead to a faithful PC surrogate. Several error metrics were computed to check the validity of the PC model.

As a follow-up research, we are seeking to infer all nine KPP parameters in a calibration where the wind is treated as an adjustable parameter. The additional number of parameters dramatically increases the number of required expensive model runs. Thus a different approach is needed where we will investigate an adaptive approach to PC (Winokur et al. 2013).

Acknowledgments. This research made use of the resources of the Supercomputing Laboratory and computer clusters at King Abdullah University of Science and Technology (KAUST) in Thuwal, Saudi Arabia. SZ and OK are supported in part by the US Department of Energy, Office of Advance Scientific Computing Research, under Award number de-sc0008789.

References

Adcroft, A. J., 1995: Numerical algorithms for use in a dynamical model of the ocean. Ph.D. thesis, University of London, 1-117 pp., London, England.

- Alexanderian, A., O. Le Maitre, H. N. Najm, M. Iskandarani, and O. M. Knio, 2011: Multiscale Stochastic Preconditioners in Non-intrusive Spectral Projection. *Journal of Scientific Computing*, 1–35, doi:10.1007/s10915-011-9486-2, URL <http://dx.doi.org/10.1007/s10915-011-9486-2>.
- Alexanderian, A., J. Winokur, I. Sraj, A. Srinivasan, M. Iskandarani, W. C. Thacker, and O. M. Knio, 2012: Global sensitivity analysis in an ocean general circulation model: A sparse spectral projection approach. *Computational Geosciences*, **16** (3), 757–778.
- Babuška, I., F. Nobile, and R. Tempone, 2007: A stochastic collocation method for elliptic partial differential equations with random input data. *SIAM J. Numer. Anal.*, **45** (3), 1005–1034.
- Berg, E. v., and M. P. Friedlander, 2007: SPGL1: A solver for large-scale sparse reconstruction. URL <http://www.cs.ubc.ca/labs/scl/index.php/Main/Spgl1>.
- Berveiller, M., B. Sudret, and M. Lemaire, 2006: Stochastic finite element: a non intrusive approach by regression. *Eur. J. Comput. Mech.*, **15**, 81–92.
- Blatman, G., and B. Sudret, 2011: Adaptive Sparse Polynomial Chaos Expansion Based on Least Angle Regression. *J. Comput. Phys.*, **230** (6), 2345–2367.
- Conrad, P. R., and Y. Marzouk, 2013: Adaptive Smolyak Speudospectral Approximations. *SIAM J. Sci. Comp.*, **35** (6), 2643–2670.
- Constantine, P. G., M. S. Eldred, and E. T. Phipps, 2012: Sparse pseudospectral approximation method. *Computer Methods in Applied Mechanics and Engineering*, **229-232**, 1–12.
- Crestaux, T., O. Le Maitre, and J.-M. Martinez, 2009: Polynomial chaos expansion for sensitivity analysis. *Reliability Engineering & System Safety*, **94** (7), 1161–1172.

- Donoho, D., 2006: Compressed sensing. *IEEE Trans. Inform. Theory*, **52** (4), 1289–1306.
- Doostan, A., and H. Owhadi, 2011: A non-adapted sparse approximation of {PDEs} with stochastic inputs. *Journal of Computational Physics*, **230** (8), 3015 – 3034, doi:http://dx.doi.org/10.1016/j.jcp.2011.01.002, URL <http://www.sciencedirect.com/science/article/pii/S0021999111000106>.
- Elsheikh, A. H., I. Hoteit, and M. F. Wheeler, 2014: Efficient Bayesian inference of subsurface flow models using nested sampling and sparse polynomial chaos surrogates. *Computer Methods in Applied Mechanics and Engineering*, **269** (0), 515–537, doi:http://dx.doi.org/10.1016/j.cma.2013.11.001, URL <http://www.sciencedirect.com/science/article/pii/S004578251300296X>.
- Ferreira, D., and J. Marshall, 2006: Formulation and implementation of a residual-mean ocean circulation model. *Ocean Modelling*, **13** (1), 86–107, doi:http://dx.doi.org/10.1016/j.ocemod.2005.12.001, URL <http://www.sciencedirect.com/science/article/pii/S1463500305001022>.
- Forget, G., 2010: Mapping ocean observations in a dynamical framework: a 2004-2006 ocean atlas. *Journal of Physical Oceanography*, **40**, 1201–1221.
- Ge, L., and K. Cheung, 2011: Spectral Sampling Method for Uncertainty Propagation in Long-Wave Runup Modeling. *Journal of Hydraulic Engineering*, **137** (3), 277–288, doi:10.1061/(ASCE)HY.1943-7900.0000301, URL [http://dx.doi.org/10.1061/\(ASCE\)HY.1943-7900.0000301](http://dx.doi.org/10.1061/(ASCE)HY.1943-7900.0000301).
- Gelman, A., 2006: Prior distributions for variance parameters in hierarchical models (comment on article by Browne and Draper). *Bayesian Analysis*, **1** (3), 515–534, doi:10.1214/06-BA117A, URL <http://dx.doi.org/10.1214/06-BA117A>.

- Gelman, A., J. B. Carlin, H. L. Stern, and D. B. Rubin, 2004: *Bayesian Data Analysis*. 2nd ed., Chapman and Hall/CRC, 668 pp.
- Gerstner, T., and M. Griebel, 2003: Dimension-adaptive tensor-product quadrature. *COMPUTING*, **71** (1), 65–87, doi:10.1007/s00607-003-0015-5.
- Ghanem, R. G., and P. D. Spanos, 1991: *Stochastic Finite Elements: A Spectral Approach*. Springer-Verlag, New York, 214 pp.
- Haario, H., E. Saksman, and J. Tamminen, 2001: An Adaptive Metropolis Algorithm. *Bernoulli*, **7** (2), 223–242, URL <http://www.jstor.org/stable/3318737>.
- Hoteit, I., B. Cornuelle, and P. Heimbach, 2010: An eddy-permitting, dynamically consistent adjoint-based assimilation system for the tropical Pacific: Hindcast experiments in 2000. *Journal of Geophysical Research*, **115**(C3), C03 001.
- Hoteit, I., B. Cornuelle, V. Thierry, and D. Stammer, 2008: Impact of resolution and optimized ECCO forcing on simulation of the tropical Pacific. *Journal of Atmospheric and Oceanic Technology*, **25**, 131–147.
- Jackson, C., and G. Huerta, 2015: Empirical bayes approach to scaling uncertainties in the representation of climate data covariances. *to be submitted to Geoscientific Model Development*.
- Jackson, C., M. Sen, and P. Stoffa, 2004: An efficient stochastic bayesian approach to optimal parameter and uncertainty estimation for climate model predictions. *Journal of Climate*, **17**(14), 2828–2841.
- Large, W. G., G. Danabasoglu, S. C. Doney, and J. C. McWilliams, 1997: Sensitivity to Surface Forcing and Boundary Layer Mixing in a Global Ocean Model: Annual-Mean Climatology. *Journal of Physical Oceanography*, **27** (11), 2418–2447.

- Large, W. G., and P. R. Gent, 1999: Validation of vertical mixing in an equatorial ocean model using large eddy simulations and observations. *Journal of Physical Oceanography*, **29**, 449–464.
- Large, W. G., J. C. McWilliams, and S. C. Doney, 1994a: Oceanic vertical mixing: a review and a model with a non-local boundary layer parameterization. *Reviews in Geophysics*, **32**, 363–403.
- Large, W. G., J. C. McWilliams, and S. C. Doney, 1994b: Oceanic vertical mixing: A review and a model with a non-local boundary layer parameterization. *Reviews of Geophysics*, **32**, 363–403.
- Le Maître, O. P., and O. M. Knio, 2010: *Spectral Methods for Uncertainty Quantification*. Springer-Verlag.
- Li, G., M. Iskandarani, M. Le Henaff, J. Winokur, O. Le Maitre, and O. Knio, 2015: Quantifying initial and wind forcing uncertainties in the Gulf of Mexico . *Submitted*.
- Lin, G., and G. E. Karniadakis, 2009: Sensitivity analysis and stochastic simulations of non-equilibrium plasma flow. *International Journal for Numerical Methods in Engineering*, **80** (6-7), 738–766, doi:10.1002/nme.2582.
- Malinverno, A., 2002: Parsimonious Bayesian Markov chain Monte Carlo inversion in a nonlinear geophysical problem. *Geophysical Journal International*, **151** (3), 675–688, doi:10.1046/j.1365-246X.2002.01847.x, URL <http://dx.doi.org/10.1046/j.1365-246X.2002.01847.x>.
- Marshall, J., A. Adcroft, C. Hill, L. Perelman, and C. Heisey, 1997a: A finite-volume, incompressible Navier Stokes model for studies of the ocean on parallel computers. *Journal of Geophysical Research*, **102**, 5753–5766.
- Marshall, J., C. Hill, L. Perelman, and A. Adcroft, 1997b: Hydrostatic, quasi-hydrostatic, and non-hydrostatic ocean modeling. *Journal of Geophysical Research*, **102**, 5733–5752.

- Marzouk, Y. M., and H. N. Najm, 2009: Dimensionality reduction and polynomial chaos acceleration of Bayesian inference in inverse problems. *Journal of Computational Physics*, **228** (6), 1862–1902, doi:10.1016/j.jcp.2008.11.024.
- Marzouk, Y. M., H. N. Najm, and L. A. Rahn, 2007: Stochastic spectral methods for efficient Bayesian solution of inverse problems. *Journal of Computational Physics*, **224** (2), 560–586, doi:10.1016/j.jcp.2006.10.010.
- Mattern, P., K. Fennel, and M. Dowd, 2012: Estimating time-dependent parameters for a biological ocean model using an emulator approach. *Journal of Marine Systems*, **96–97**, 32–47, doi:http://dx.doi.org/10.1016/j.jmarsys.2012.01.015.
- McPhaden, M. J., and Coauthors, 1998: The Tropical Ocean-Global Atmosphere (TOGA) observing system: A decade of progress. *Journal of Geophysical Research*, **103**, 14 169–14 240.
- Nobile, F., R. Tempone, and C. G. Webster, 2008: A sparse grid stochastic collocation method for partial differential equations with random input data. *SIAM J. Numer. Anal.*, **46** (5), 2411–2442.
- Olson, R., R. Sriver, M. Goes, N. M. Urban, H. D. Matthews, M. Haran, and K. Keller, 2012: A climate sensitivity estimate using Bayesian fusion of instrumental observations and an Earth System model. *Journal of Geophysical Research*, **117**, D04 103.
- Parzen, E., 1962: On Estimation of a Probability Density Function and Mode. *The Annals of Mathematical Statistics*, **33** (3), 1065–1076, URL <http://www.jstor.org/stable/2237880>.
- Peng, J., J. Hampton, and A. Doostan, 2014: A weighted l_1 minimization approach for sparse polynomial chaos expansions. *Journal of Computational Physics*, **267**, 92–111.
- Petras, K., 2000: On the Smolyak Cubature Error for Analytic Functions. *Advances in Computational Mathematics*, **12**, 71–93.

- Petras, K., 2001: Fast calculation in the Smolyak algorithm. *Num. Algo.*, **26**, 93–109.
- Petras, K., 2003: Smolyak cubature of given polynomial degree with few nodes for increasing dimension. *Numer. Math.*, **93** (4), 729–753, doi:10.1007/s002110200401.
- Reagan, M. T., H. N. Najm, R. G. Ghanem, and O. M. Knio, 2003: Uncertainty Quantification in Reacting Flow Simulations Through Non-Intrusive Spectral Projection. *Combustion and Flame*, **132**, 545–555.
- Roberts, G. O., and J. S. Rosenthal, 2009: Examples of Adaptive {MCMC}. *Journal of Computational and Graphical Statistics*, **18** (2), 349–367, doi:10.1198/jcgs.2009.06134.
- Silverman, B. W., 1986: *Density estimation: for statistics and data analysis*. Chapman and Hall, London.
- Sivia, D. S., 2006: *Data Analysis - A Bayesian Tutorial*. Oxford Science Publications, Oxford, 264 pp.
- Smolyak, S. A., 1963: Quadrature and interpolation formulas for tensor products of certain classes of functions. *Dokl. Akad. Nauk SSSR*, **4**, 240–243.
- Socolofsky, S., T. Bhaumik, and D. Seol, 2008: Double-plume integral models for near-field mixing in multiphase plumes. *Journal of Hydraulic Engineering*, **134** (6), 772–783.
- Sraj, I., M. Iskandarani, A. Srinivasan, W. C. Thacker, and O. M. Knio, 2013a: Drag Parameter Estimation using Gradients and Hessian from a Polynomial Chaos Model Surrogate. *Monthly Weather Review*, **142**, 933–941, doi:10.1175/MWR-D-13-00087.1.
- Sraj, I., K. Mandli, O. M. Knio, and I. Hoteit, 2014: Uncertainty Quantification and Inference of Manning’s Friction Coefficients using DART Buoy Data during the Tohoku Tsunami. *Ocean Modelling*, **83**, 82–97, doi:10.1016/j.ocemod.2014.09.001.

- Sraj, I., and Coauthors, 2013b: Bayesian Inference of Drag Parameters using Fanapi AXBT Data. *Monthly Weather Review*, **141**, 2347–2367, doi:10.1175/MWR-D-12-00228.1.
- Sudret, B., 2008: Global sensitivity analysis using polynomial chaos expansions. *Reliability Engineering & System Safety*, **93** (7), 964–979, doi:DOI:10.1016/j.ress.2007.04.002.
- Tarantola, A., 2005: *Inverse Problem Theory and Methods for Model Parameter Estimation*. SIAM, Philadelphia, PA, 342 pp.
- van den Berg, E., and M. P. Friedlander, 2008: Probing the pareto frontier for basis pursuit solutions. *SIAM Journal on Scientific Computing*, **31** (2), 890–912, doi:10.1137/080714488, URL <http://link.aip.org/link/?SCE/31/890>.
- Villegas, M., F. Augustin, A. Gilg, A. Hmadi, and U. Wever, 2012: Application of the Polynomial Chaos Expansion to the simulation of chemical reactors with uncertainties. *Mathematics and Computers in Simulation*, **82** (5), 805–817, doi:10.1016/j.matcom.2011.12.001.
- Wagman, B., C. Jackson, F. Yao, S. Zedler, and I. Hoteit, 2014: Metric of the 2-6 day sea-surface temperature response to wind stress in the Tropical Pacific and its sensitivity to the K-Profile-Parameterization of vertical mixing. *Ocean Modelling*, **79**, 54–64.
- Wang, J., and N. Zabaras, 2005: Hierarchical Bayesian models for inverse problems in heat conduction. *Inverse Problems*, **21** (1), 183, URL <http://stacks.iop.org/0266-5611/21/i=1/a=012>.
- Wang, S., M. Iskandarani, A. Srinivasan, C. Thacker, J. Winokur, and O. Knio, 2015: Propagation of uncertainty and sensitivity analysis in an integral oil-gas plume model . *submitted*.
- Winokur, J., P. Conrad, I. Sraj, O. Knio, A. Srinivasan, W. Thacker, Y. Marzouk, and M. Iskandarani, 2013: A priori testing of sparse adaptive polynomial chaos expansions using an

ocean general circulation model database. *Computational Geosciences*, **17** (6), 899–911, doi: 10.1007/s10596-013-9361-3, URL <http://dx.doi.org/10.1007/s10596-013-9361-3>.

Xiu, D., and J. S. Hesthaven, 2005: High-order collocation methods for differential equations with random inputs. *SIAM J. Sci. Comput.*, **27** (3), 1118–1139.

Xiu, D., and D. M. Tartakovsky, 2004: Uncertainty quantification for flow in highly heterogeneous porous media. *Computational Methods in Water Resources: Volume 1*, W. G. G. Cass T. Miller Matthew W. Farthing, and G. F. Pinder, Eds., Developments in Water Science, Vol. 55, Part 1, Elsevier, 695–703.

Zedler, S., G. Kanschä, R. Körtz, and I. Hoteit, 2012: A new approach for the determination of the drag coefficient from the upper ocean response to a tropical cyclone: a feasibility study. *Journal of Oceanography*, **68** (2), 227–241.

Zedler, S. E., C. Jackson, F. Yao, A. S. Heimbach, P. and I. Hoteit: A new kind of test statistics for parameter estimation within noisy data and uncertain forcing: Application to the K-Profile Parameterization of ocean boundary layering mixing. *in revision*.

LIST OF TABLES

Table 1. List of KPP parameters to be estimated using Bayesian inference along with their MITGCM default values and assumed uniform prior. 35

Table 2. Reporting mean and variance of test statistic in addition to sensitivity index of each uncertain parameter. 36

Parameter	Parameter Name	Symbol	Default Value	Uniform Prior $[a, b]$
p_1	Critical Bulk Richardson number	Ri_c	0.3	$[0.1, 1.0]$
p_2	Critical gradient Richardson number	Ri_g	0.7	$[0.1, 1.0]$
p_3	Structure function, unstable forcing, momentum	$\phi_{m,unst}$	16	$[3.60, 331.06]$
p_4	Structure function, unstable forcing, tracer	$\phi_{s,unst}$	16	$[7.77, 67.02]$
p_5	Nonlocal transport	C^*	10.0	$[5.0, 15.0]$

1 TABLE 1. List of KPP parameters to be estimated using Bayesian inference along with their MITGCM default
2 values and assumed uniform prior.

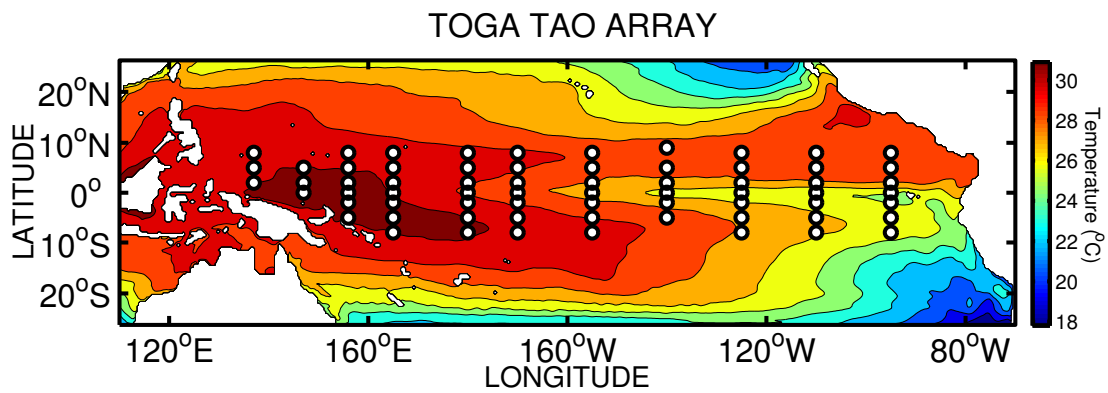
μ_E	σ_E	T_1	T_2	T_3	T_4	T_5
325.62	30.97	0.6907	0.1450	0.2079	0.0472	0.0784

TABLE 2. Reporting mean and variance of test statistic in addition to sensitivity index of each uncertain parameter.

LIST OF FIGURES

Fig. 1.	Model domain showing 2004-2007 averaged sea surface temperature using default MIT-GCM KPP parameters and the TOGA/TAO mooring array.	39
Fig. 2.	Projection of the Smolyak Sparse quadrature nodes corresponding to level 5 on 2D plane.	40
Fig. 3.	Test statistic (E) versus KPP parameters at the 903 sparse quadrature nodes. Each panel plots E against one the uncertain parameters as indicated. The red dashed line in each panel corresponds to the case when the other parameters are fixed to the midpoint of their uniform prior value i.e. $\xi_i = 0$	41
Fig. 4.	PC expansion normalized coefficients $ e_k/e_0 $ for PC order up to $r = 5$. The dashed vertical lines separate the PC expansion terms into degrees. The coefficients were calculated using NISP.	42
Fig. 5.	Comparing test statistic (E) from MITGCM model runs superimposed with their PC surrogate counterparts constructed using NISP. The shown cases correspond to the sparse quadrature.	43
Fig. 6.	Test statistic (E) from MITGCM model runs when varying Ri_c only infinitesimally	44
Fig. 7.	PC expansion normalized coefficients $ e_k/e_0 $ for PC order up to $r = 5$. The dashed vertical lines separate the PC expansion terms into degrees. The coefficients were calculated using BPDN.	45
Fig. 8.	<i>pdfs</i> of test statistic E with increasing order of PC constructed using BPDN-estimated PC surrogate model.	46
Fig. 9.	Comparing test statistic (E) from MITGCM model runs with their PC surrogate counterparts (Left) Superimposed (Right) Scatter plot. The shown cases correspond to the sparse quadrature and PC is constructed using BPDN. The normalized relative error (NRE) is also indicated.	47
Fig. 10.	Projection of the independent random sample on the $\xi_1 - \xi_2$ plane.	48
Fig. 11.	Test statistic (E) versus KPP parameters at the 954 independent random sample. Each panel plots E against one the uncertain parameters as indicated.	49
Fig. 12.	Comparing test statistic (E) from MITGCM model runs with their PC surrogate counterparts (Left) Superimposed (Right) Scatter plot. The shown cases correspond to the independent random sample and PC is constructed using BPDN. The relative normalized error (NRE) is also indicated.	50
Fig. 13.	2D Smolyak Sparse quadrature nodes corresponding to levels 1, 2, 3 and 4 in the canonical vector space.	51
Fig. 14.	Normalized relative error using different number of nodes corresponding to different Smolyak levels of refinement.	52
Fig. 15.	Response curves of test statistic (E) function of different parameters (in canonical space). For each curve, the other four parameters are set to $\xi_j = 0$	53

Fig. 16.	Response surfaces of test statistic (E) function of (top) Ri_c and Ri_g , (bottom) Ri_c versus $\phi_{m,unst}$ (in canonical space). For each surface, the other three parameters are set to $\xi_j = 0$.	54
Fig. 17.	Chains of parameters using MCMC sample from PC surrogate constructed. MITGCM default values are indicated as horizontal lines.	55
Fig. 18.	Running mean of the different chains.	56
Fig. 19.	$pdfs$ of parameters using KDE	57



5 FIG. 1. Model domain showing 2004-2007 averaged sea surface temperature using default MITGCM KPP
 6 parameters and the TOGA/TAO mooring array.

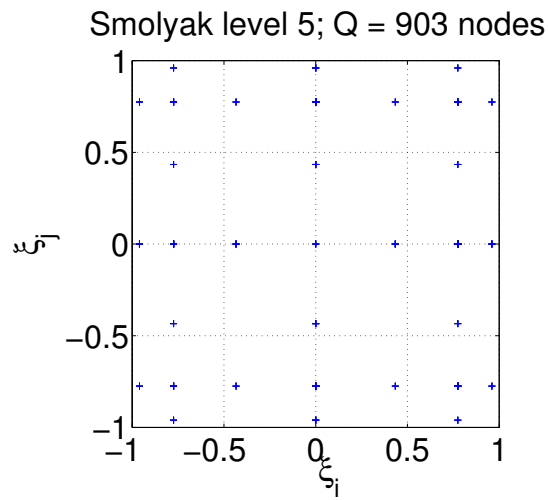


FIG. 2. Projection of the Smolyak Sparse quadrature nodes corresponding to level 5 on 2D plane.

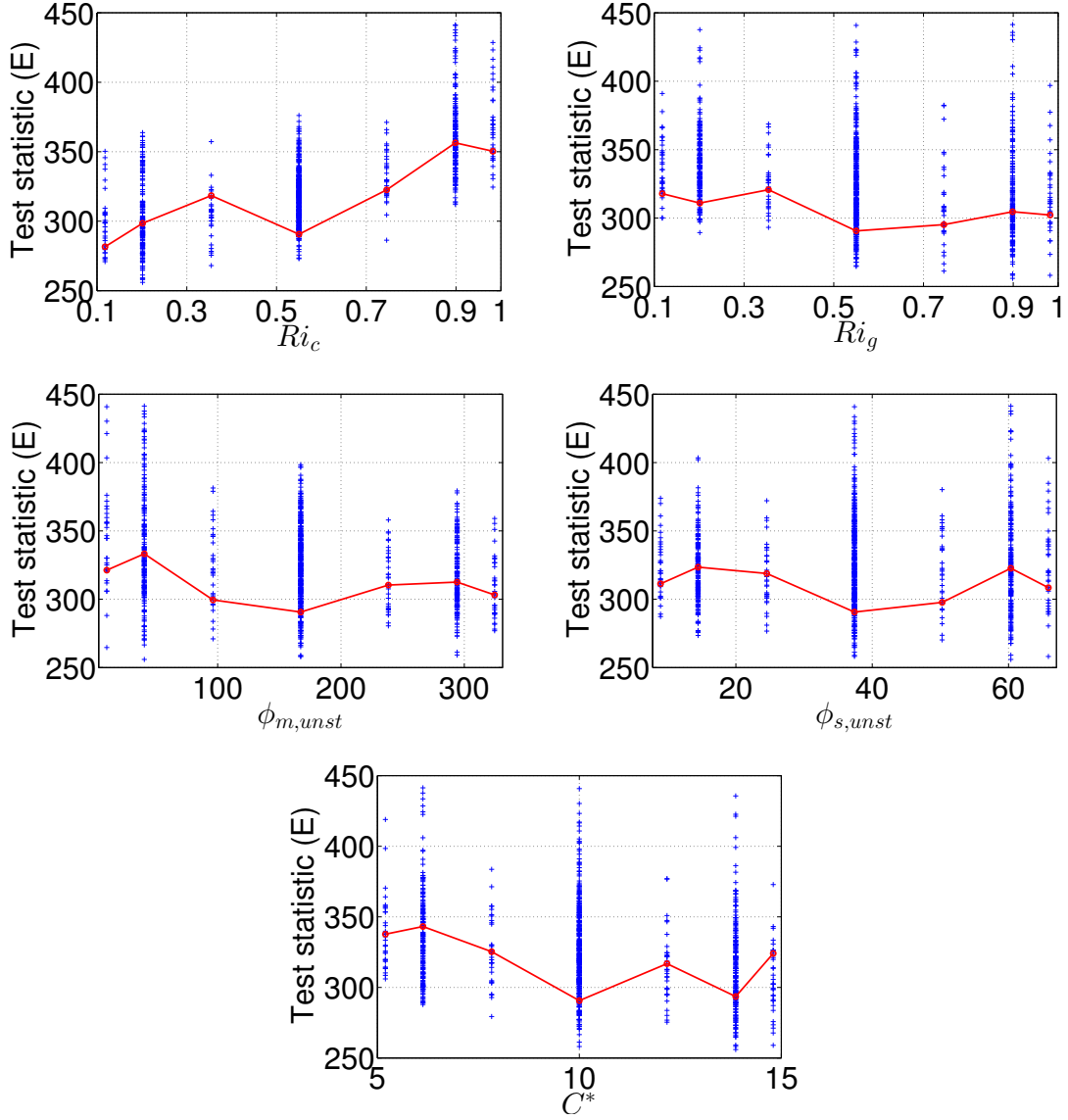


FIG. 3. Test statistic (E) versus KPP parameters at the 903 sparse quadrature nodes. Each panel plots E against one of the uncertain parameters as indicated. The red dashed line in each panel corresponds to the case when the other parameters are fixed to the midpoint of their uniform prior value i.e. $\xi_i = 0$.

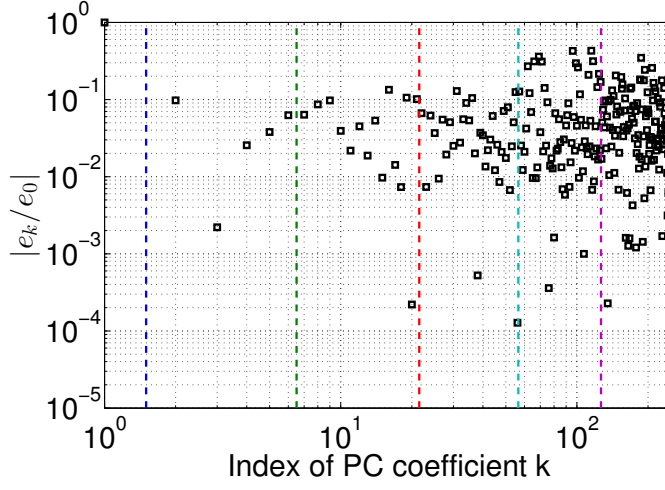


FIG. 4. PC expansion normalized coefficients $|e_k/e_0|$ for PC order up to $r = 5$. The dashed vertical lines separate the PC expansion terms into degrees. The coefficients were calculated using NISP.

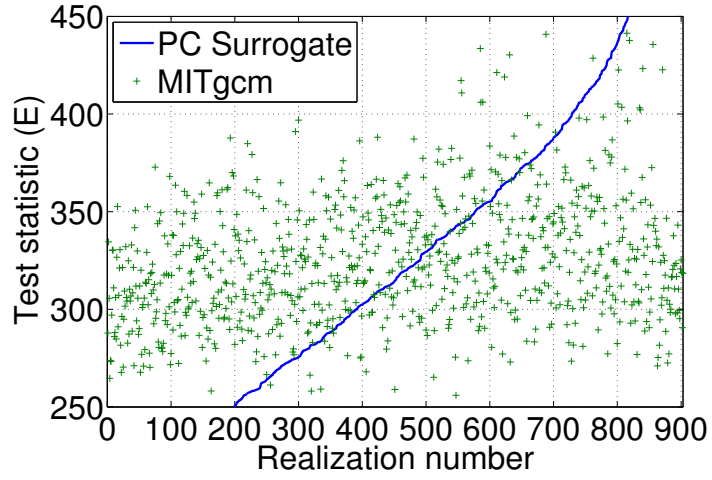


FIG. 5. Comparing test statistic (E) from MITGCM model runs superimposed with their PC surrogate counterparts constructed using NISP. The shown cases correspond to the sparse quadrature.

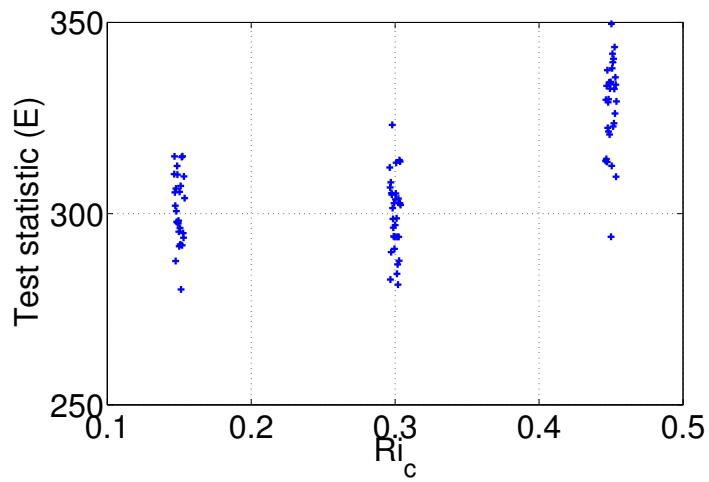


FIG. 6. Test statistic (E) from MITGCM model runs when varying Ri_c only infinitesimally

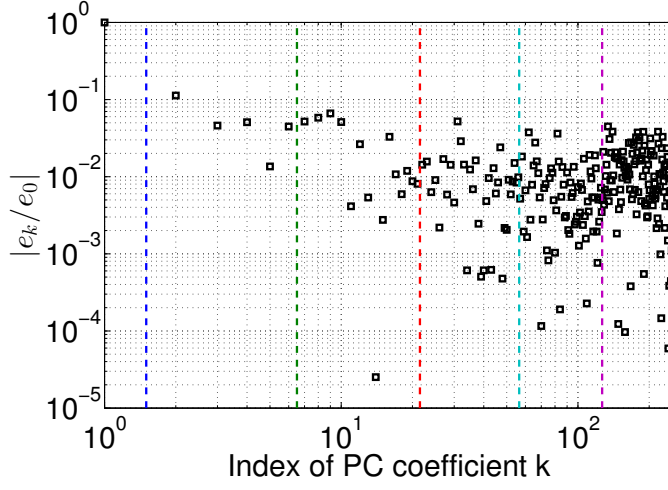


FIG. 7. PC expansion normalized coefficients $|e_k/e_0|$ for PC order up to $r = 5$. The dashed vertical lines separate the PC expansion terms into degrees. The coefficients were calculated using BPDN.

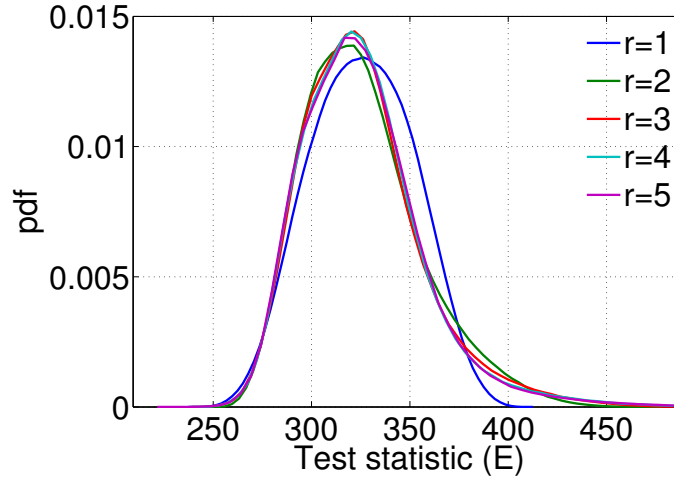


FIG. 8. *pdfs* of test statistic E with increasing order of PC constructed using BPDN-estimated PC surrogate model.

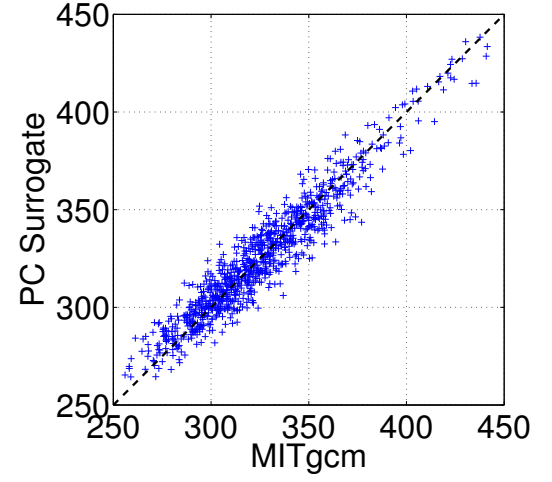
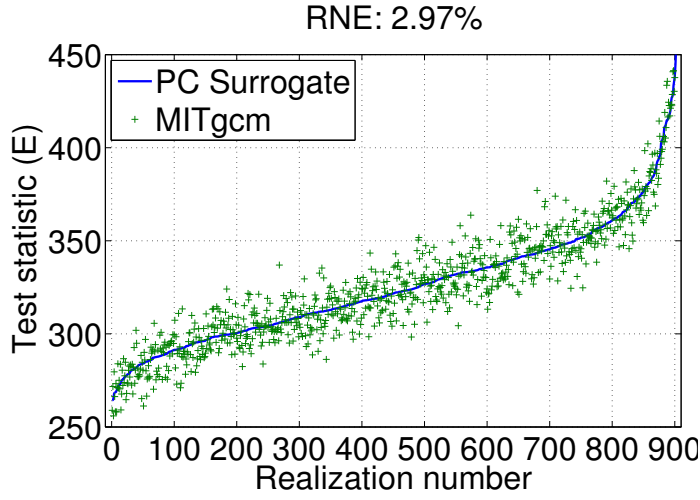


FIG. 9. Comparing test statistic (E) from MITGCM model runs with their PC surrogate counterparts (Left) Superimposed (Right) Scatter plot. The shown cases correspond to the sparse quadrature and PC is constructed using BPDN. The normalized relative error (NRE) is also indicated.

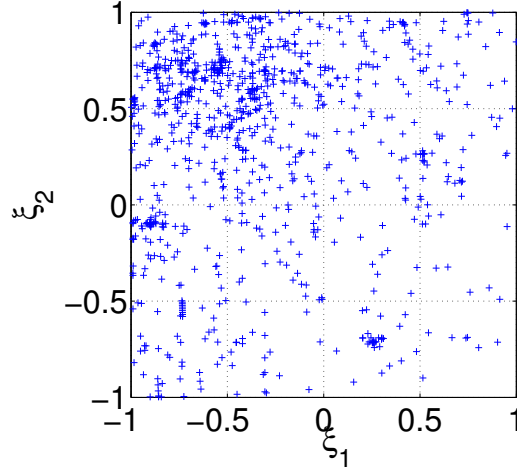
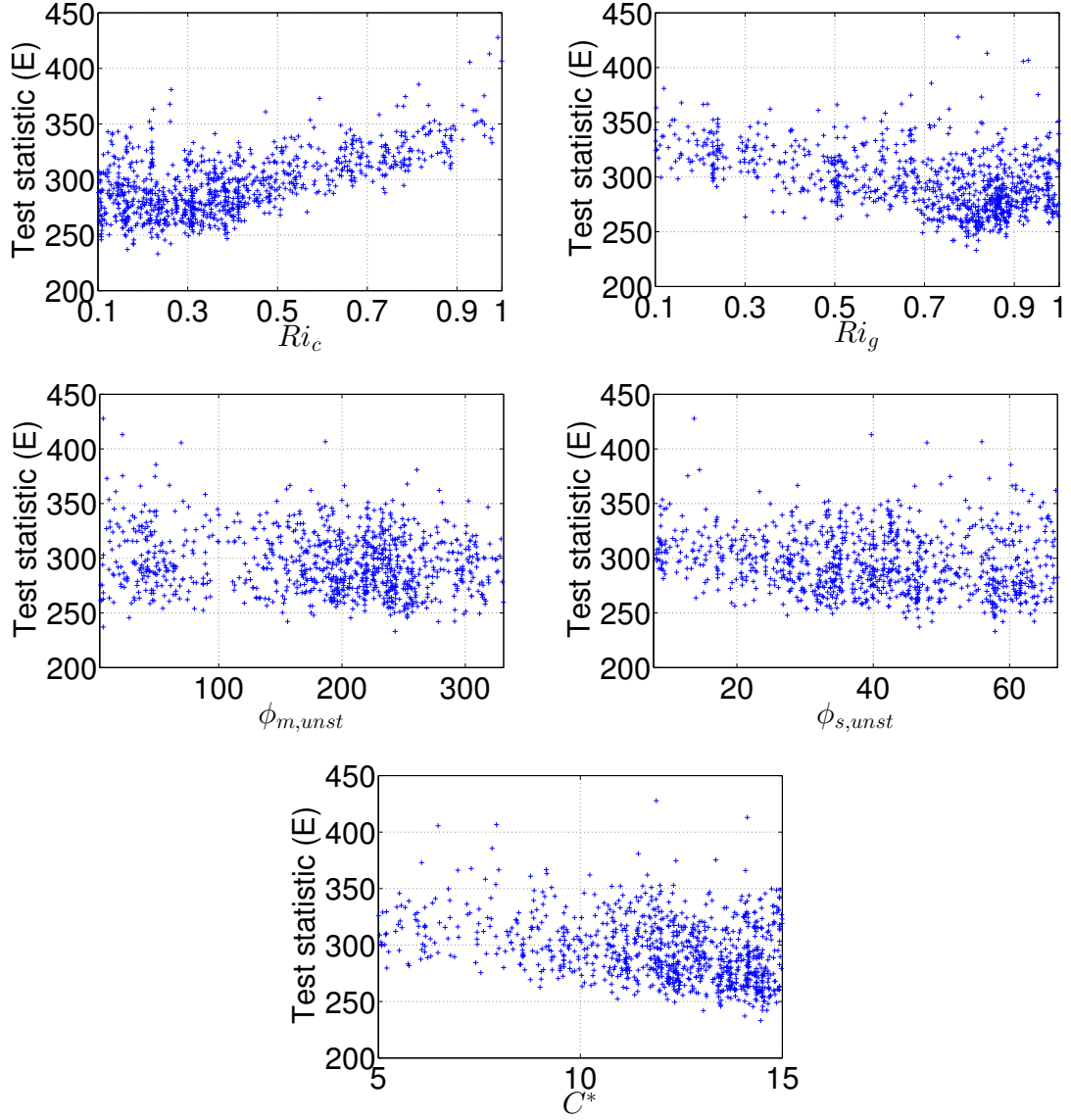


FIG. 10. Projection of the independent random sample on the $\xi_1 - \xi_2$ plane.



21 FIG. 11. Test statistic (E) versus KPP parameters at the 954 independent random sample. Each panel plots E
 22 against one the uncertain parameters as indicated.

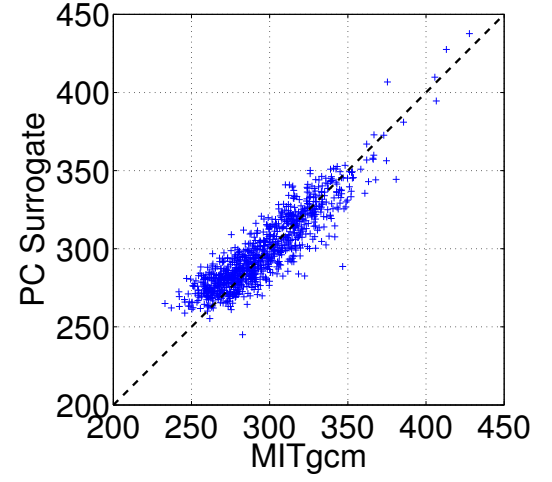
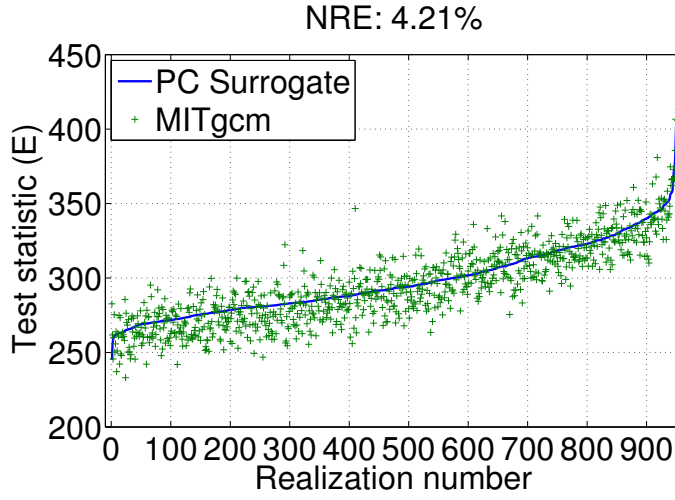


FIG. 12. Comparing test statistic (E) from MITGCM model runs with their PC surrogate counterparts (Left) Superimposed (Right) Scatter plot. The shown cases correspond to the independent random sample and PC is constructed using BPDN. The relative normalized error (NRE) is also indicated.

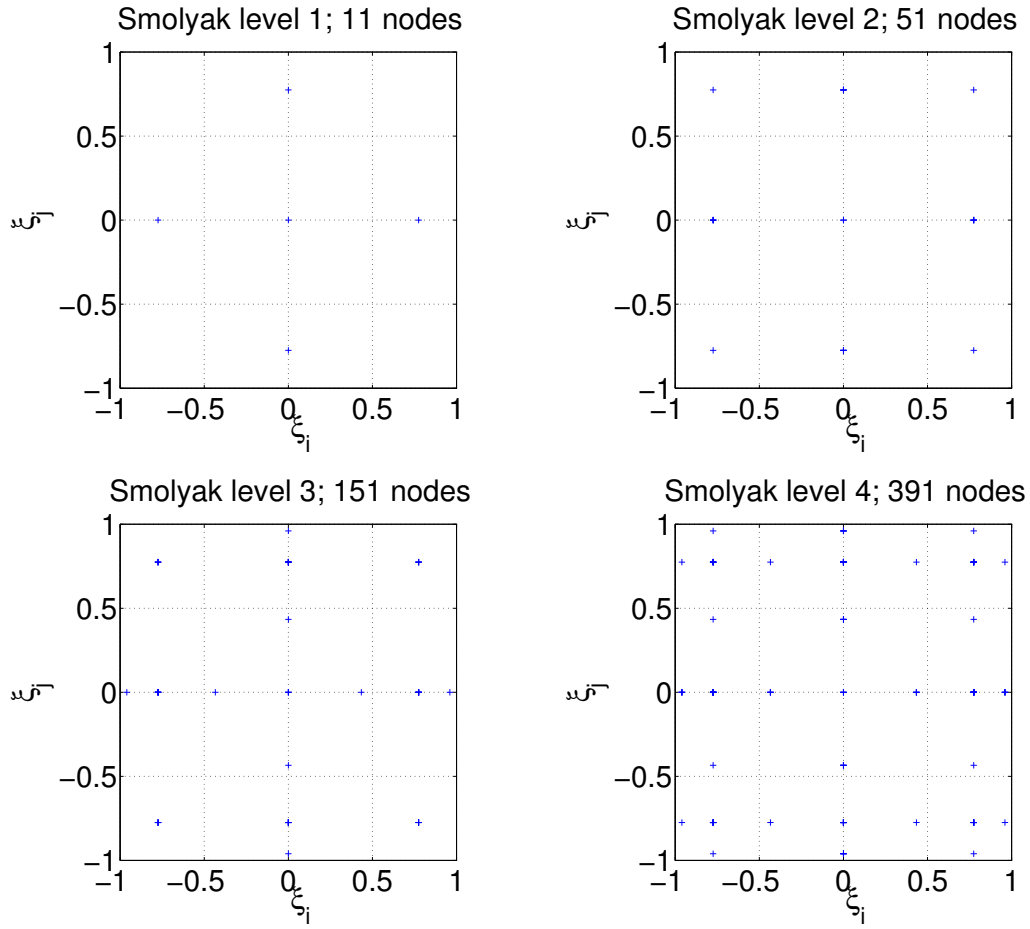
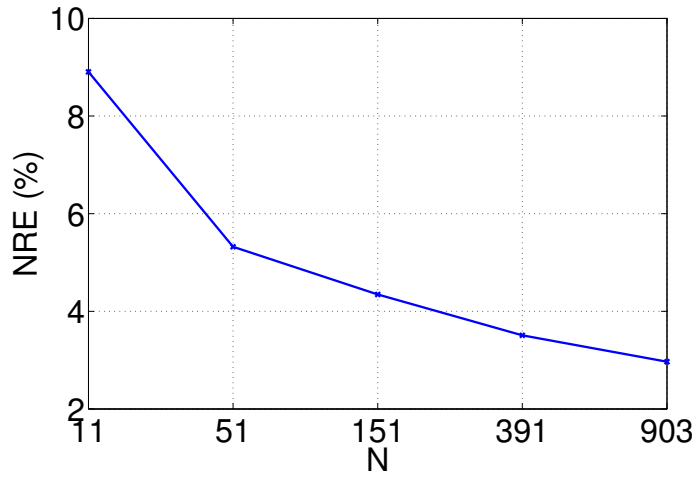


FIG. 13. 2D Smolyak Sparse quadrature nodes corresponding to levels 1, 2, 3 and 4 in the canonical vector space.



28 FIG. 14. Normalized relative error using different number of nodes corresponding to different Smolyak levels
29 of refinement.

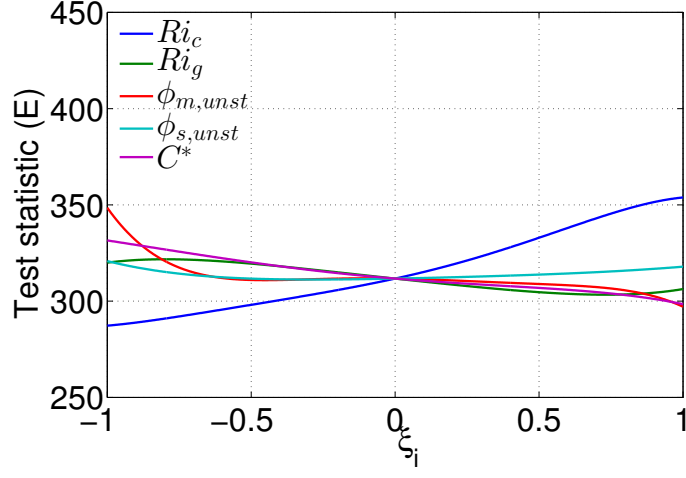


FIG. 15. Response curves of test statistic (E) function of different parameters (in canonical space). For each curve, the other four parameters are set to $\xi_j = 0$.

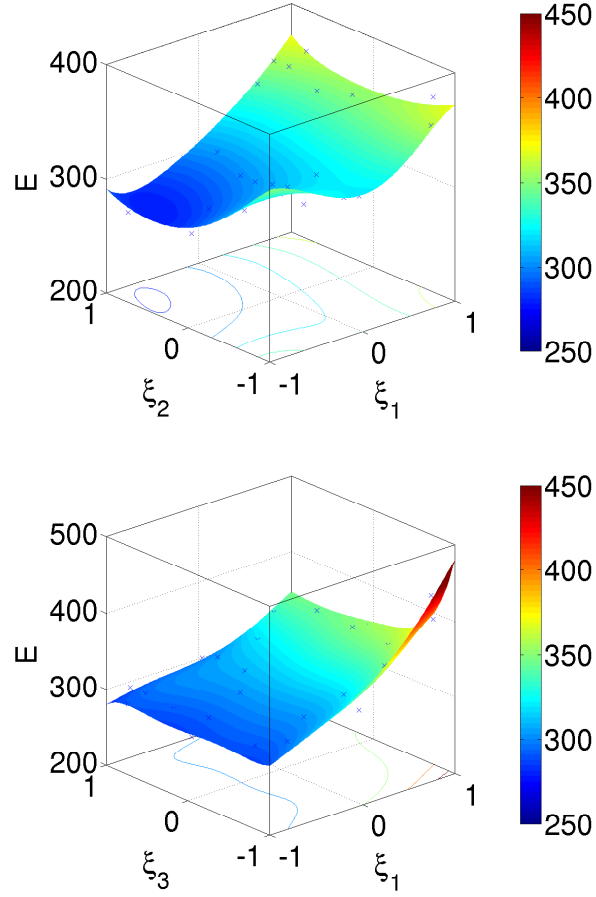


FIG. 16. Response surfaces of test statistic (E) function of (top) Ri_c and Ri_g , (bottom) Ri_c versus $\phi_{m,unst}$ (in canonical space). For each surface, the other three parameters are set to $\xi_j = 0$.

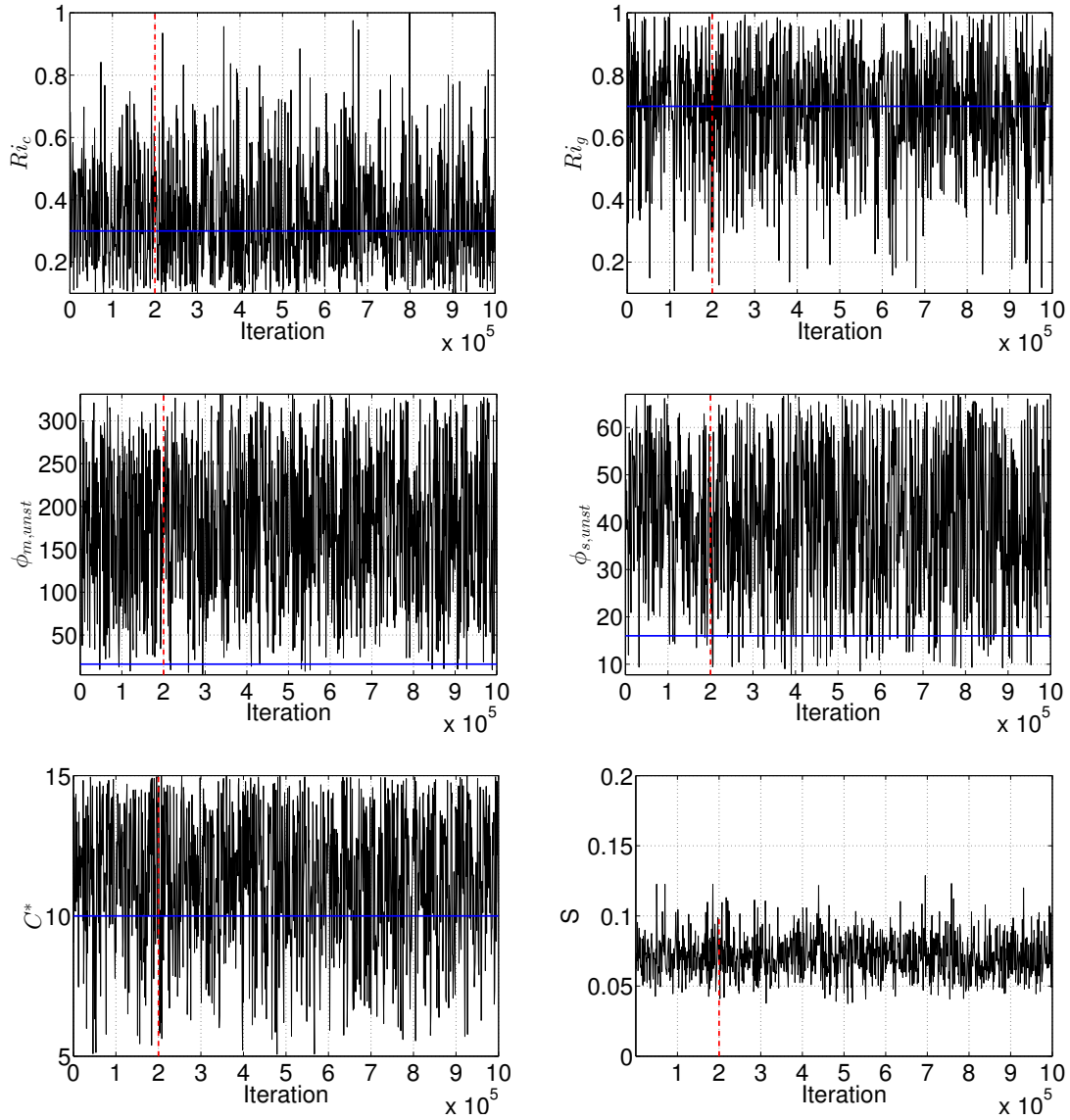


FIG. 17. Chains of parameters using MCMC sample from PC surrogate constructed. MITGCM default values are indicated as horizontal lines.

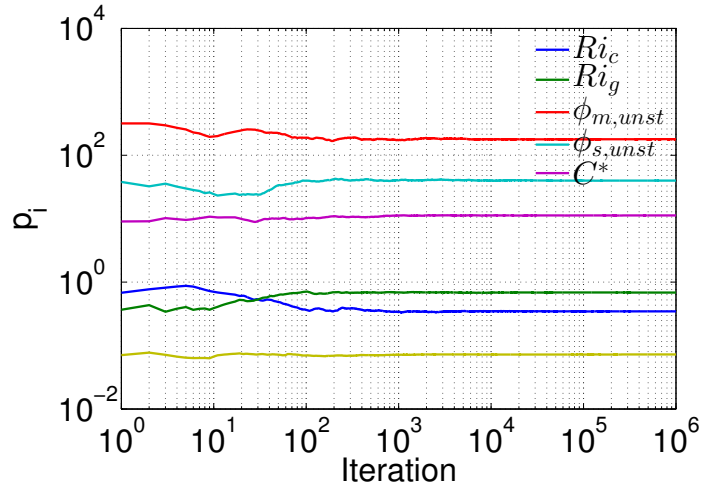


FIG. 18. Running mean of the different chains.

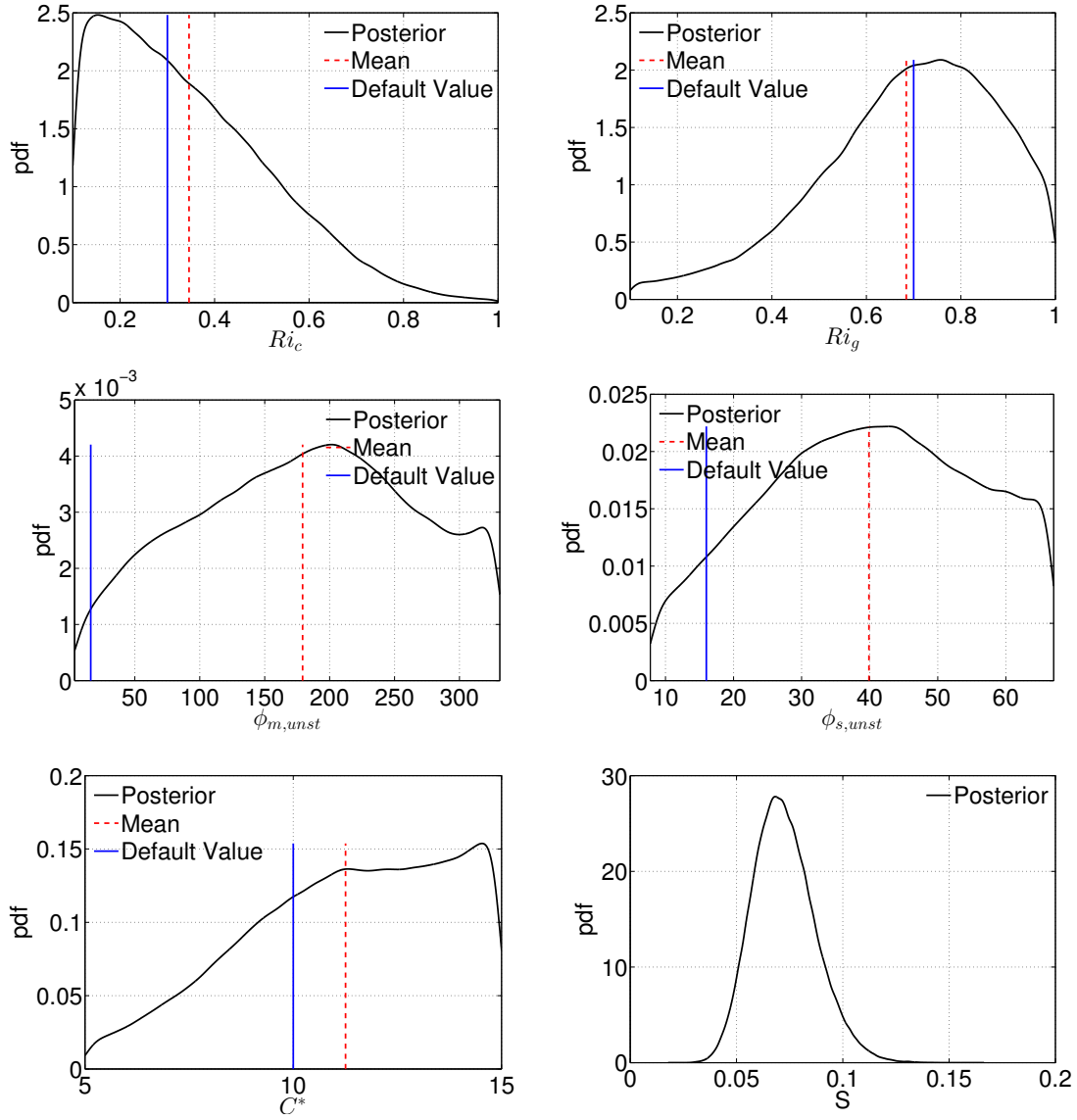


FIG. 19. *pdfs* of parameters using KDE .